L'épistémologie bayésienne

Baptiste Le Bihan

http://www.baptistelebihan.org

BA2b Introduction à la philosophie des sciences

Remerciements: Christian Wüthrich, Pablo Carnino, Augustin Baas

Plan

La connaissance incertaine

- Les sciences empiriques ne donnent que rarement des connaissances entièrement certaines.
- Il semble difficile d'avoir entièrement confiance dans le raisonnement inductif.
- La confiance dans les raisonnements déductifs est conditionnée par notre confiance dans la vérité des prémisses de ces raisonnements.
- Les énoncés d'observations sont incertains : marges d'erreur dans les mesures.
- La nature elle-même semble parfois se comporter de manière indéterminée fondamentalement (désintégration radioactive), ou dérivée (raisonnement statistique).

$\Rightarrow$ Pour toutes ces raisons, il nous faut appréhender l'incertitude en sciences.

Épistémologie/Théorie de la confirmation bayésienne

- Un des développements les plus importants dans l'épistémologie du 20e siècle.
- Elle offre un modèle formel et mathématique rigoureux qui lie la croyance en certaines hypothèses et leur confirmation ou leur infirmation par des données.
- Le modèle est probabiliste: il assigne des probabilités aux croyances.
- Idée générale: une donnée à l'appui e confirme une hypothèse h dans le cas où elle augmente la probabilité de h, i.e., $P(h|e) > P(h)$.
- Les probabilités doivent être «mises à jour» de la façon prédite par le théorème de Bayes, de sorte que le degré de mise à jour de la croyance en une hypothèse soit la probabilité de l'hypothèse conditionnellement aux données.

$\Rightarrow$ Les lois du calcul des probabilités sont des contraintes sur des degrés rationnels de croyance (ou de confiance).

Axiomes de Kolmogorov (axiomes des probabilités)

Donnée: classe $\mathcal{S}$ de propositions $a, b, c, \ldots$

Introduire une fonction de probabilité $P(\cdot)$ sur $\mathcal{S}$ qui associe $\mathcal{S}$ à l'intervale clos $[0, 1]$ telle que les axiomes suivants soient valables:

Axiome (1: non-négativité)

$P(x) \geq 0$ *pour tout* x *dans* $\mathcal{S}$*; i.e. toutes les probabilités sont non-négatives.*

Axiome (2: unité de mesure)

$P(x) = 1$ *si* x *dans* $\mathcal{S}$ *est une tautologie; i.e. si* x *est une proposition qui est vraie dans tous les cas possibles, alors elle a la probabilité 1.*

Axiome (3: additivité)

Pout tout x *et* y *dans* $\mathcal{S}$*, si* x *et* y *sont des propositions mutuellement exclusives, alors* $P(x \vee y) = P(x) + P(y)$*.*

Probabilité conditionnelle

Définition (Probabilité conditionnelle)

Étant donné une fonction de probabilité $P(x)$ comme définie sur la diapositive précédente, la probabilité conditionnelle $P(s|t)$ de s étant donné t est définie comme

$$P(s|t) := \frac{P(s\&t)}{P(t)}.$$

Exemple: Jeu de cartes de jass

$$\begin{aligned} P(\heartsuit|\text{rouge}) &= \frac{P(\heartsuit\&\text{rouge})}{P(\text{rouge})} = \frac{1/4}{1/2} = \frac{1}{2} \\ P(\text{roi}|\heartsuit) &= \frac{P(\text{roi}\&\heartsuit)}{P(\heartsuit)} = \frac{1/36}{1/4} = \frac{1}{9} \end{aligned}$$

Mise à jour bayésienne des croyances

Les bayésiens font (une version plus compliquée de) la présupposition épistémologique suivante:

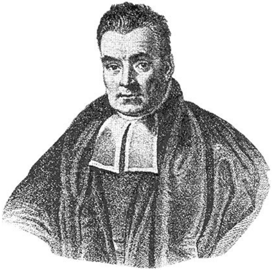
Principe (Conditionnalisation)

À partir des probabilités initiales (antérieures) $P_i(h)$ de n'importe quel hypothèse h et ayant acquis de nouvelles données tel qu'on a observé les données e (c'est-à-dire $P(e) = 1$), la rationalité dicte que l'on mette à jour ses probabilités initiales afin d'obtenir ses probabilités finales (postérieures) pour «conditionner» sur e

$$P_i(h) \to P_f(h) = P_i(h|e).$$

Donc, on doit trouver un moyen de calculer $P_i(h|e)$ comme une fonction de $P_i(h)$.

Théorème du Révérend Thomas Bayes (1702-1761)



Pour toute proposition h et e, on a

$$P(h|e) = \frac{P(h) \cdot P(e|h)}{P(e)}$$
$$= \frac{P(h) \cdot P(e|h)}{P(e|h) \cdot P(h) + P(e|\neg h) \cdot P(\neg h)}$$

où

- $P(h)$: probabilité antérieure (*a priori*) de h
- $P(h|e)$: probabilité postérieure (*a posteriori*) de h sachant e (ou encore: de h sous condition e)
- $P(e|h)$: fonction de vraisemblance de h (pour un e connu)

Dérivation du théorème de Bayes

(a) Dérivation de la première forme (en utilisant deux fois la définition de la probabilité conditionnelle):

$$P(h|e) = \frac{P(h\&e)}{P(e)} = \frac{P(e\&h)}{P(e)} = \frac{P(e|h)\cdot P(h)}{P(e)}$$

(b) Dérivation de la forme alternative du dénominateur (en utilisant la «formule des probabilités totales» et la définition de la probabilité conditionnelle):

$$\begin{aligned} P(e) &= P(e\&h) + P(e\&\neg h) \\ &= \frac{P(e\&h)\cdot P(h)}{P(h)} + \frac{P(e\&\neg h)\cdot P(\neg h)}{P(\neg h)} \\ &= P(e|h)\cdot P(h) + P(e|\neg h)\cdot P(\neg h) \end{aligned}$$

Mise à jour bayésienne

- D'abord, déterminez la probabilité antérieure de h et les chances que e_1 soit observé étant donné h.
- Déterminez la probabilité d'observer e_1 indépendamment de h.
- Si e_1 est observé, calculez la probabilité postérieure $P(h|e_1)$ via le théorème de Bayes.
- Considérez cette probabilité postérieure comme votre nouvelle probabilité antérieure de h.
- Considérez la probabilité d'une nouvelle donnée à l'appui de e_2 et ses chances à la lumière de h.
- Si e_2 est observé, calculez la nouvelle probabilité postérieure de h via le théorème de Bayes.
- ...

Exemple 1: de quelle boîte vient le cookie?

Deux boîtes de cookies:

1. $Boîte_1$ a 10 cookies au chocolat et 30 cookies nature
2. $Boîte_2$ a 20 cookies au chocolat et 20 cookies nature

Question: Si l'on prend un cookie d'une boîte au hasard, et qu'il est nature (e), quelle est la probabilité qu'il provienne de la $Boîte_1$ (h)?

a priori: $P(h) = P(\neg h) = 0.5$

vraisemblances: $P(e|h) = 0.75$ et $P(e|\neg h) = 0.5$

Utilisez le théorème de Bayes:

$$\begin{aligned} P(h|e) &= \frac{P(h) \cdot P(e|h)}{P(e|h) \cdot P(h) + P(e|\neg h) \cdot P(\neg h)} \\ &= \frac{0.5 \cdot 0.75}{0.5 \cdot 0.75 + 0.5 \cdot 0.5} = 0.6. \end{aligned}$$

Example 2: Vient-elle/il à la fête?

Peter Godfrey-Smith (2003). *Theory and Reality: An Introduction to the Philosophy of Science*, University of Chicago Press, p. 204.

- h: hypothèse qu'elle/il est à la fête
- e: donnée selon lesquelles son vélo est parquée dehors
- $P(h)$: probabilité initiale qu'elle/il soit à la fête (avant d'avoir vu le vélo); disons que c'est 0.5
- $P(e|h)$: chances que son vélo soit parquée dehors si elle/il est à la fête; supposez que c'est 0.8
- $P(e|\neg h)$: chances que son vélo soit parquée dehors si elle/il n'est pas à la fête; supposez que c'est seulement 0.1
- $P(h|e)$: probabilité qu'elle/il soit à la fête étant donné que son vélo est parquée dehors; utilisez le théorème de Bayes:

$$P(h|e) = \frac{0.5 \cdot 0.8}{0.5 \cdot 0.8 + 0.5 \cdot 0.1} = 0.89.$$

$\Rightarrow$ voir le vélo augmente la probabilité de h de 0.5 à 0.89.

Exemple 3: dans un tribunal

Un juré doit évaluer comment des données pèsent en faveur de ou contre la culpabilité de l'accusé:

- c: hypothèse que l'accusé est coupable
- d: données indiquant que l'ADN de l'accusé correspond à l'ADN trouvée sur les lieux du crime
- $P(d|c)$: chances de voir des données d'ADN correspondantes si l'accusé est coupable; dans les crimes capitaux, typiquement très haute; ici présumée égale à 1
- $P(d|\neg c)$: chances de voir des données d'ADN correspondantes si l'accusé n'est pas coupable; très basse, on présume 1 sur un million, ou 10^{-6}
- $P(c)$: probabilité initiale que l'accusé soit coupable (a priori); dépend largement des autres preuves, circonstances, etc. Deux cas de figure: soit (A) forte suspicion antérieure ($P(c) = 0.3$), ou (B) suspicion très basse ($P(c) = 10^{-6}$)
- $P(c|d)$: probabilité que l'accusé soit coupable si l'ADN correspond; c'est ce que l'on cherche à savoir!

Cas (A): $P(c) = 0.3$

$$\begin{aligned} P(c|d) &= \frac{P(c) \cdot P(d|c)}{P(d|c) \cdot P(c) + P(d|\neg c) \cdot P(\neg c)} \\ &= \frac{0.3 \cdot 1.0}{0.3 \cdot 1.0 + 0.7 \cdot 10^{-6}} \\ &= 0.99999766667 \end{aligned}$$

Cas (B): $P(c) = 10^{-6}$

$$\begin{aligned} P(c|d) &= \frac{P(c) \cdot P(d|c)}{P(d|c) \cdot P(c) + P(d|\neg c) \cdot P(\neg c)} \\ &= \frac{10^{-6} \cdot 1.0}{10^{-6} \cdot 1.0 + (1 - 10^{-6}) \cdot 10^{-6}} \\ &\approx 0.5 \end{aligned}$$

Example 4: La recherche du *Scorpion*

- Mai 1968: Le sous-marin nucléaire américain *Scorpion* n'arrive pas à destination à Norfolk, VA.
- La US Navy est convaincue que le vaisseau s'est échoué sur les côtes de l'Est, mais des recherches approfondies ne permettent pas de retrouver l'épave.
- John Craven, expert des profondeurs à la US Navy croyait qu'il était au sud-ouest de l'archipel portuguais des *Açores* selon une triangulation d'hydrophones controversée.
- Attribution de ressources limitées (un navire) $\Rightarrow$ il les optimisât
- Craven travailla avec des mathématiciens pour optimiser la recherche, en utilisant la théorie bayésienne de la recherche.
- Octobre 1968: L'épave est retrouvée à 740 km au sud-ouest des Açores.

Théorie bayésienne de la recherche

1. La mer est divisée en carrés formant une grille.
2. Des commandants expérimentés de sous-marins sont interrogés, afin de formuler des hypothèses à propos de ce qui est arrivé au vaisseau.
3. Construction des lois de probabilité sur chaque carré correspondant à chaque hypothèse.
4. Construction de la loi de probabilité qu'on retrouve l'épave dans le carré x si elle est vraiment dans x (fonction de la profondeur).
5. Combinaison de toutes ces lois de probabilité (de 3 et 4) afin de produire une grille de l'ensemble des probabilités; elle donne la probabilité de trouver l'épave dans un carré si on fait des recherches dans ce carré (pour chaque carré).
6. Construction d'un chemin de recherche partant du carré avec la plus haute probabilité qui passe ensuite par des zones de haute probabilité, puis de probabilité intermédiaire, avant d'arriver aux zones de basse probabilité.
7. Révision constante de la loi de probabilité d'ensemble à mesure que la recherche avance, i.e. si on a cherché dans un carré sans succès, alors la probabilité que l'épave y soit diminue grandement (quoique habituellement pas nulle), et la probabilité de la trouver ailleurs doit être augmentée; cette révision suit le théorème de Bayes.

Exemple 5: Le V1@gra et les filtres bayésiens

- But: savoir les chances qu'un email soit du spam sur la base des mots qui y figurent.
- Idée de base: un filtre anti-spam bayésien.
  - Liste des mots dans les emails entrants,
  - assigne à chaque mot une probabilité qu'il apparaisse dans un email de spam (les mots mal épelés ont un score très élevé), et
  - emploie ces probabilités comme données entrantes dans la formule de Bayes afin de déterminer si un email est du spam ou non.
- Tout d'abord il faut entraîner le filtre anti-spam en lui montrant des emails de spam et de non-spam (de plus en plus automatisé).
- Le filtre enregistre tous les mots des messages d'entraînement (incluant nom de host, adresse IP, tag HTML, etc.) dans une banque de données.

⇒ Le filtre calcule la probabilité qu'un mot apparaisse dans un spam sur la base de sa fréquence dans la banque de données (la «spamicité» de chaque mot).

- Une spamicité de 0.5 est neutre, plus haute (moindre) signifie qu'il apparaît souvent dans des messages (non-)spam.
- Ensuite le filtre utilise la formule de Bayes afin de calculer la spamicité d'ensemble d'un email, sur la base de la spamicité de tous les mots qu'il contient.

⇒ le message est placé dans le filtre à spam si la spamicité est supérieure à 0.5.

- Généralement: Les filtres à spam bayésiens sont très efficaces car (1) ils s'adaptent au circonstances individuelles (la banque de donnée est construite pour chaque usager), et (2) ils apprennent avec le temps et mettent à jour la banque de données.

Qu'est-ce que la probabilité?

hasard (objectif)	vs.	degré de croyance (subjectif)
↓		↓
Comment mesurer?		Comment mesurer?
↓		↓
fréquence des occurrences		attitude de pari

Les axiomes de la probabilité doivent s'appliquer aux deux!

Bayésiannisme subjectiviste

- «Probabilité comme degré de croyance personnelle» (générée par le libre choix, la socialisation, l'évolution, etc.).
- La probabilité d'un événement est simplement la certitude avec laquelle un agent bayésien s'attend à ce que l'événement se produise.
- Idée principale: «croyance rationnelle» devrait être comprise comme une généralisation de l'attitude de pari: étant donné une certaine quantité d'information et étant demandé une prédiction, à quelle cote est-ce qu'on parierait sur sa prédiction?
- Parier sur h à une cote de $x : 1$ signifie être prêt à risquer de perdre x CHF si h s'avère fausse, contre un gain de CHF 1 si h est vraie.
- Si vos cotes subjectivement équitables pour un pari sur h sont $x : 1$, alors votre degré de croyance en h est $x/(x+1)$.
- Il est bien sûr possible que le degré subjectif de croyance viole les axiomes, mais...

Le théorème du «Dutch book»

Si le degré de croyance subjective de quelqu'un viole un axiome de Kolmogorov, alors cette personne devrait accepter une combinaison de paris qui reviennent à un «Dutch book», i.e. la combinaison de paris qui, s'ils devaient être acceptés, garantissent à la personne une perte!

Exemple simple: votre degré de croyance que le prochain pile-ou-face tombera sur «face» (f) est de 0.55, et votre degré de croyance qu'il tombera sur «pile» (p) est 0.5 $\Rightarrow$ le bookmaker peut vous proposer un ensemble de paris qui garantit que vous perdrez 5 centimes sur chaque franc parié.

Table de décision sur une combinaison de paris [en francs]:

	$f(\neg p)$	$p(\neg f)$
pariez 0.55	1	0
pariez 0.50	0	1
payez 1.05 au total	gagnez 1	gagnez 1

La solution bayésienne au paradoxe «vleu» («noanc»)

- Supposez qu'on vous présente deux arguments inductifs à partir du même ensemble d'observations des corbeaux noirs, l'un argumentant que tous les corbeaux sont noirs, l'autre qu'ils sont noancs.
- Pourquoi une des deux inductions est-elle meilleure que l'autre?
- Réponse bayésienne standard: les deux sont OK, mais la plupart des gens assignerait une probabilité antérieure plus haute à l'hypothèse «noir» qu'à la «noanc».
- Réaction: vrai, cela montre une différence, mais cela explique-t-il pourquoi l'hypothèse «noanc» à une probabilité antérieure plus basse?
- Le bayésiannisme n'offre aucune critique de la décision subjective d'assigner une probabilité antérieure plus haute à l'hypothèse «noanc», pour autant que les probabilités soient cohérentes de manière interne et mises à jour comme il faut.

(1) Problème des antérieures

- L'ensemble initial des probabilités antérieures peut être choisi librement (excepté pour 0 et 1).
- Mais comment peut-on critiquer une assignation étrange des probabilités antérieures, tant qu'elle suit les axiomes?
- Réponse bayésienne: cela n'importe pas car l'ensemble initial des probabilités antérieures est évacué asymptotiquement (convergence, théorème de l'estimation stable).
- Problème technique: conversement, on a aussi pour toute quantité de données, et toute mesure d'accord, un ensemble de probabilités antérieures telles que ces données ne vont pas finir par mettre les deux personnes d'accord (Henry Kyburg).
- Les conditions pour les théorèmes doivent être en place: par exemple, il doit y avoir un accord concernant les vraisemblances $P(e_i|h)$ et la pertinence de certaines données particulières.
- Les présuppositions des théorèmes ne s'appliquent même pas un petit peu dans des contextes réalistes scientifiquement.

(2) Explication et truismes méthodologiques (Glymour)

- Une théorie de la confirmation satisfaisante doit être à même d'expliquer des truismes généraux méthodologiques ainsi que des jugements particuliers qui ont eu lieu dans l'histoire des sciences.
- Le bayésiannisme ne peut pas expliquer:
  - pourquoi une hypothèse est ad hoc,
  - pourquoi un ensemble de données est plus varié qu'un autre,
  - pourquoi on devrait préférer des données à l'appui plus variées,
  - pourquoi on devrait préférer les théories plus simples (e.g. problème de «curve-fitting» – ajuster la courbe).
- Généralement: Le bayésiannisme ne peut pas expliquer pourquoi on fait les projections qu'on fait et pourquoi des données diverses peuvent être pertinentes de manière différente (et pourquoi on pourrait être en désaccord sur l'attribution de cette pertinence).

(3) Problème de l'ancienne donnée (Glymour)

- Problème de l'ancienne donnée: D'anciennes données peuvent en fait confirmer une nouvelle théorie, mais d'après la cinématique bayésienne elles ne le peuvent pas.
- Supposez que e est connu avant que l'hypothèse h ne soit introduite au temps t.
- Parce que e est connu à t, $P_t(e) = 1$.

$\Rightarrow$ la vraisemblance de h est aussi 1: $P_t(e|h) = 1$.

$$P_t(h|e) = \frac{P_t(h) \cdot P_t(e|h)}{P_t(e)} = P_t(h)$$

$\Rightarrow$ La probabilité postérieure de h est la même que sa probabilité antérieure!

En guise de conclusion

Alan Hájek and Stephan Hartmann (2010). Bayesian epistemology. In J. Dancy et al. (eds.), *A Companion to Epistemology*, Blackwell, 93-105.

Alan Hájek et Stephan Hartmann contrastent deux conceptions de l'épistémologie bayésienne:

«Selon un certain point de vue, il ne peut pas y avoir d'épistémologie bayésienne: le bayésiannisme ne rend pas justice aux aspects essentiels de la connaissance et de la croyance et, de ce fait, il n'offre pas authentiquement une épistémologie. Selon un autre point de vue, le bayésiannisme doit supplanter l'épistémologie traditionnelle: alors que cette dernière s'est épuisées dans des débats sans fin sur le scepticisme et la gettierologie, le bayésiannisme offre à l'épistémologie un programme de recherche. Nous proposons une conception plus modérée, à savoir que le bayésiannisme permet d'éclairer divers problèmes de longue date de l'épistémologie, sans pour autant les traiter tous; et bien que le bayésiannisme ouvre de nouvelles aires de recherche fascinantes, il ne met aucunement fin aux préoccupations centrales de l'épistémologie traditionnelle.» (93, trad. Baptiste Le Bihan)